

日本語従属節の意味分類基準策定について：「鳥バンク」節間意味分類体系再構築の提案

著者	松本 理美
雑誌名	国立国語研究所論集
号	15
ページ	107-133
発行年	2018-07
URL	http://doi.org/10.15084/00001599

日本語従属節の意味分類基準策定について

——「鳥バンク」節間意味分類体系再構築の提案——

松本理美

立命館大学大学院 博士課程／国立国語研究所 共同研究員

要旨

本研究は、日本語研究のための日本語従属節の意味分類基準の策定のために、「鳥バンク」節間意味分類体系の再構築の検討を行う。「鳥バンク」は日英対訳コーパス中の文型パターンの分析と分類を重ね、日英機械翻訳のための意味分類体系の構築を行っている。「鳥バンク」節間意味分類体系を基準とした『現代日本語書き言葉均衡コーパス』に対する従属節アノテーションを行ったところ、意味分類のタグ付与に作業員間の齟齬が生じた。その齟齬の分析により、「鳥バンク」節間意味分類体系を日本語従属節の意味分類に転用することの問題点が明らかになった。価値ある研究成果である「鳥バンク」節間意味分類体系を日本語研究に有効活用するための修正を試みる。

従属節のアノテーション作業の齟齬分析では、副詞節と並列節の間に生じた従属節の機能レベルでの齟齬が顕著であった。従属節の中には並列節と副詞節の機能を共有している節があり、意味分類体系の再構築においては第1段階を4分類から3分類に修正することを提案する。副詞節の下位では、文脈や発話意図が節の種類の特定に影響することから、意味分類の細かさを調整し、アノテーションの精度が保てる意味分類基準策定の可能性を示した。また、従属節のアノテーションの手掛かりとなる「節間キーワード」の問題（複数の項目に同形の「節間キーワード」が存在すること、複数の「節間キーワード」が複合した別の「節間キーワード」が存在することなど）を整理し、高精度アノテーションのための従属節意味分類基準案を提示した。

節境界の認定や節の意味分類にはいくつかの解決しがたい問題が存在するが、「鳥バンク」節間意味分類体系を基準とした従属節の意味分類アノテーション作業における問題点を議論し、「鳥バンク」節間意味分類体系を再構築することにより、日本語研究のための安定した日本語従属節の意味分類基準策定の可能性を示す*。

キーワード: 「鳥バンク」、日本語従属節、節の意味分類、節アノテーション

1. はじめに

最近のコーパス開発の発展に伴い、日本語学、日本語教育などにおいてもコーパスを利用した研究や教材開発などが多く見られる。例えば、現代日本語の節連鎖構造について話し言葉と書き言葉を比較した丸山（2014）は、コーパスを利用して量的な側面から文法にアプローチした研究を行っている。堀・江田（2013）では複数のコーパスからデータを抽出し、日本語教育での利用を目的とした「日本語文法項目用例文データベース『はごろも』」（堀・江田 2013）について述

* 本稿は国立国語研究所コーパス開発センターの共同研究プロジェクト「コーパスアノテーションの拡張・統合・自動化に関する基礎研究」（プロジェクトリーダー：浅原正幸、2016–2021 年度）の成果の一部である。また、本研究の一部は JSPS 科研費（課題番号：15K12888、研究代表者：浅原正幸）の助成を受けている。

なお、本稿は言語資源活用ワークショップ 2017（2017 年 9 月 5 日、国立国語研究所）で発表した内容に加筆・修正したものである。発表に際し、多くの方々に貴重な御意見・御助言を賜った。また本稿執筆にあたり、プロジェクトリーダーの浅原正幸准教授、立命館大学の有田節子教授には、多くの有益な御指摘・御助言と御鞭撻をいただいた。ここに記してお礼申し上げる。

べている。

言語資源をより有効に利用するために、コーパスへのアノテーションの研究も進展しているが、日本語の節に対するアノテーション精度を保つことは、困難な課題の1つである。丸山他 (2016) にも述べられている通り、日本語学における節の認定や意味分類に関する研究は多く、長い歴史があり、従属節¹の分類からして諸説がある。意味分類²基準やアノテーション基準の策定は容易ではないが、コーパスを利用した日本語従属節の研究において、漏れなく従属節を検出し、分析データの信頼性を確保するためには、一定の基準に則った高精度のアノテーションが求められることは言うまでもない。

本研究の目的は、日本語－英語機械翻訳を主とした言語変換³のために構築されたデータバンク（以下、「鳥バンク」とする。池原 2007a）の節間意味分類体系を、日本語研究のための日本語従属節意味分類基準に再構築するための枠組みを提案することである。

『現代日本語書き言葉均衡コーパス』（以下、BCCWJ。Maekawa et al. 2014）に対する節へのアノテーション作業の中で、筆者は「鳥バンク」節間意味分類体系を基準とした日本語従属節への意味分類タグ付与作業（以下、「鳥バンク」基準のアノテーション作業）を行った。作業の基準とした「鳥バンク」節間意味分類体系は、大量の文型パターンをグループ化することにより作成されたもので、日本語の主節・従属節間の意味を 222 項目に分類し、632 の「節間キーワード」（2.1 節で説明）と日本語の文例を各項目に配した非常に価値の高い研究成果である。「鳥バンク」基準のアノテーション作業の最終段階では、作業者間・作業間の 9 割のタグ付与に一致が見られたことから、「鳥バンク」節間意味分類体系は日本語従属節の分類に有用であると考えられる。しかし、最終的に残存したタグ付与の齟齬を分析すると、英語－日本語機械翻訳に対応するよう工学的視点から開発された「鳥バンク」節間意味分類体系を日本語従属節の意味分類に転用したことに起因する齟齬が大半を占めることが確認された（3 節で詳説）。日本語学の視点から「鳥バンク」節間意味分類体系を修正し、日本語研究のための従属節意味分類基準を策定することはアノテーションの精度を保つために有効である。

本稿では、2 節で「鳥バンク」の詳細および日本語コーパスと日本語従属節の意味分類に関連する先行研究について述べ、3.1 節で、筆者が作業者として関わった「鳥バンク」基準のアノテーション作業について述べる。3.2 節では、意味分類アノテーション作業で生じたタグ付与の齟齬の分析に基づき、「鳥バンク」節間意味分類体系を日本語従属節の意味分類に転用することの問題点について実例を示しながら議論したうえで、「鳥バンク」節間意味分類体系についての修正案を示す。タグ付与の齟齬についての議論と修正提案は以下の順で述べる：従属節の機能分類に

¹ 本研究では主節以外の節を全て従属節とする。

² 池原 (2009) は「節の意味分類」について、「節は単一の事象を表現したもので、その内容は単文に相当するため、節の意味分類は単文の意味分類に従う」(p. 265) とし、節が表現している事象についての意味分類を指している。本研究で着目している従属節の意味分類を池原 (2009) では「節間の意味分類」としているが、本稿では「従属節の意味分類」と呼ぶことにする。

³ 日本語の重文と複文を対象に、文や節を構成している語の言い換えだけでは、文や節全体の意味を言い換えられないような表現の機械翻訳を目指している。

における齟齬と分類階層の修正 (3.2.1 節), 補足節の下位分類における齟齬と分類項目の修正 (3.2.2 節), 名詞修飾節の下位分類における齟齬と分類項目の修正 (3.2.3 節), 副詞節の下位分類における齟齬と分類項目の修正 (3.2.4 節), 並列節の下位分類における齟齬と分類項目の修正 (3.2.5 節), 「節間キーワード」の問題点と修正 (3.2.6 節) の順である。4 節で, 各従属節における意味分類タグ付与の齟齬の分析と修正案を踏まえて提案した, 「鳥バンク」節間意味分類体系再構築による, 日本語研究のための日本語従属節意味分類基準の枠組みの全容を示す。5 節で本研究のまとめと今後の課題について述べる。

2. 関連研究

昨今のコーパス開発の急進により, 機械学習に基づく言語解析器の開発, アノテーションの自動化などが進んでいるが, 日本語研究や日本語教育研究に関するコーパス開発において, 日本語学と工学の協働の歴史はまだ浅い。コーパスに対するアノテーションの精度を保つためには, 節のアノテーション基準・意味分類基準を策定することが求められる。日本語の節の認定, 節の意味の特定には日本語学的に判断が難しい問題もあるが, 日本語学と工学の協働による基準策定が必須の課題であると言える。

本節では, 日本語従属節に関し, 言語学と工学を発展的に結合させ, 困難な課題に挑む先行研究について述べる。まず, 従属節の意味分類自動化の先駆的研究であり, 本研究の礎でもある「鳥バンク」について池原 (2007a, 2007b, 2009) を基に, 節間意味分類体系に着目して概観する。次に日本語従属節の意味分類についての重要な関連研究として丸山他 (2004, 2016), 佐藤・丸山 (2015), 佐藤他 (2016) について述べる。

2.1 「鳥バンク」について

池原 (2007a, 2009) によると, 「鳥バンク」は日本語の重文と複文を対象とした, 大規模な「意味類型パターン辞書」(約 22.7 万件) を収録したデータバンク⁴である。「重文・複文の持つ統語的な構造と意味の関係に着目」(池原 2009: 254) し, 粗いレベルでの従属節の意味分類の半自動化の可能性を追求している。意味的等価変換方式⁵の実現を目指したもので, これに必要な重文・複文の意味が, 日本語表現の意味が英語表現で定義されている文型パターンによって体系的に分類されている。池原 (2009: 257) は「日本語表現の意味分類の粒度として, 英語表現との意味的対応が取れる範囲以上の物は不要である」としており, 「鳥バンク」節間意味分類体系は, 英語への機械翻訳を目的に構築されたもの⁶であることが確認できる。

「鳥バンク」の意味分類体系の 1 つである, 主節と従属節の意味的関係を分類した「節間意味

⁴ 池原悟氏が研究代表者を務めた科学技術振興機構 JST の戦略的創造研究事業 CREST の研究成果をインターネット上のホームページ「鳥バンク」で公開している。

⁵ 「概念の持つ構造と表現形式の関係に着目したもので, 原言語の文型パターンと目的言語の文型パターンそれぞれを意味類型化 (意味的に分類) し, 両言語間で同一の意味を持つ文型パターングループ間で写像を行う方法である。」(池原 2009: 253)

⁶ 同じ言語内での言い換えも含んでいる。

分類体系」について簡潔に述べる。「鳥バンク」節間意味分類体系は従属節が4段階に分類されており、第1段階から第2段階へと段階が進むほど分類は詳細になる。「鳥バンク」節間意味分類体系の構成について分類の着目点、分類項目数を表1に示した。

表1 「鳥バンク」節間意味分類体系の構成

段階		分類の着目点	分類項目数
第1段階	大分類	統語的特徴	4
第2段階	中分類	統語的色彩の強いレベル	27
第3段階	小分類 (1)	節間キーワード	37
第4段階	小分類 (2)	意味による用例分類	154

第1段階は、益岡・田窪 (1992)、益岡 (1997) を参考に、従属節を「補足節 (補語相当節)」「名詞修飾節 (連体節)」「副詞節 (副詞的連用節)」「並列節 (並列型連用節)」に4分類した「統語的特徴を考慮した大分類」である。第2段階は、それを「統語的色彩の強いレベル」でさらに27分類した中分類である。次に第3段階を「節間キーワードを考慮した小分類 (1)」として、「約1,000件の文型パターンを対象に従属節と主節とを連結するキーワードとその意味に着目した用例分析を行い、分類を詳細化」(池原 2009: 259) して37の下位分類を行っている。最後は第4段階の「用例分類過程での小分類 (2)」である。第3段階までで「得られた分類体系を12万件の文型パターン辞書 (単語レベル) に適用し、文型パターンを実際にグループ化した」(池原 2009: 265) 分類項目の新設や細分化により154の下位分類を行い、「最終的に222種類からなる意味分類体系を作成」(池原 2009: 265) している。

「鳥バンク」節間意味分類体系の概要として、第1段階 (大分類) から第3段階 (小分類 (1)) までの分類項目と、各項目の「節間キーワード」の例を表2に示した。これが本研究の基となるものである。

表2 「節間意味分類体系」(概略)

節間意味分類			節間キーワード
大	中	小分類 (1)	※代表例のみ
補足節 (補語相当節)			
	名詞節		
		コト型	ことを, ことに
		ノ型	のを, というの (が は)
		トコロ型	ところ, ところを, ところは
		節+格助詞型	節+に, 節+も
疑問節			
		選択疑問文	〜か〜か, か, かどうか
		疑問語疑問文	疑問語+か, 疑問語+のか
引用節			
		直接引用	」と
		間接引用	と, ように, なんて
名詞修飾節 (連体節)			
	補足語修飾節		
		限定的	連体形+名詞
		非限定的	連体形+固有名詞
内容節			連体形+思考・発言名詞
縮約形修飾節			連体形+因果名詞
機能的表現			
		形式名詞修飾節	連体形+形式名詞
		文末表現相当	連体形+名詞+述部
		慣用的表現	連体形+名詞+述部
		副詞節相当	連体形+名詞+で
用言+接続表現+の			ての, てからの, ながらの
その他			連体形+名詞, 節+の+名詞
副詞節 (副詞的連用節)			
	時		
		事態の時	時 (に) (は), てから
		事態の継続期間	時から, まで (は), にわたって
	因果関係		
		原因	ので, 故, 事で
		結果	(た)(が)為(に), た事から, 結果
	理由根拠		から, を以て, 事から

節間意味分類			節間キーワード
大	中	小分類 (1)	※代表例のみ
	条件・譲歩		
		法則的	ば
		偶有的	と, たら, たところ
		仮想的	としたら, とすると
		反事实的	たら〜 (だろう でしょう)
		譲歩	ても
	付帯状況・様態		
		付帯状況	つつ, ながら, まま
		様態	通り (に), 風に, 様 (に) (は)
	逆接		が, けれど, のに
	目的		(が) 為 (に), べく, 様
程度		くらい, ほど (に)	
前提		からには, 以上は, 限り	
手段		事で, 事により, 事によって	
二者関係		一方で, より (も), のと同様	
相関		とともに, に連れ	
判断, 主観		とは, ようでは, からして	
場面		ところでは, ところによると	
限定		だけで, を除き, 以外 (に)	
独立		感情動詞+事には, 〜ものか, にも	
その他			
		助動詞相当表現	ざるを
		慣用的表現	A と言ったら A, をいいことに
		副詞相当表現	節+接続表現
		発言・思考内容	たく, とて
2文			、
並列節 (並列型連用節)			
	順接的並列		
		総記の並列	し, 連用中止, て
		例示の並列	ば, たり, や
		累加の並列	だけではなく, 上に
		平行	とともに, と同時に, 且つ
		否定の並列	なくて, ないで
		選択	か
逆接的並列		が, けれど, けれども	

※池原 (2007b) 「意味類型パターン記述言語仕様書」の表4 (p.17) に基づき, 一部に修正を加えた。

「鳥バンク」節間意味分類体系では, それぞれの分類項目に 632 の「節間キーワード」を配し, その一つ一つに「節間キーワード」を含んだ日本語の文例を示している (日本語の文例等を含む「鳥バンク」節間意味分類体系は池原 (2007b: 付録4) を参照されたい)。「節間キーワード」とは, 「従属節と主節とを連結するキーワード」(池原 2009: 259) であり, 「節間キーワード」となる語の種類を従属節の種類ごとに定義している。例えば, 補足節の「節間キーワード」は, 名詞節では形式名詞「こと」「の」「ところ」など, 疑問節は「か」「かどうか」など, 引用節では「と」などの引用節を導く表現などである。名詞修飾節では被修飾名詞も含めた「連体形+名詞」を「節

間キーワード」とし、副詞節、並列節では従属節と主節の意味的な関係を決める接続表現部分を「節間キーワード」としている。

2.2 現代日本語の従属節意味分類に関連する最近の研究

近年、コーパス構築、開発の急進により、日本語学と工学両分野の協働による様々な日本語解析システムが進化の過程にあり、語や文字だけでなく、複雑な文法についても解析器によるアンテーション自動化システムの整備が各分野で、あるいは分野を越えて進められている。

本節では、現代日本語の従属節の意味分類に関連する最新の研究について述べる。

丸山他（2004）は、形態素解析された日本語テキストから「節境界（clause boundary）」の位置と種類を自動的に検出する「節境界検出プログラム CBAP（Clause Boundary Annotation Program）」の開発について述べている。益岡・田窪（1992）に記述されている節境界の表現をベースに、形態素の接続による節境界検出ルールを人手により作成している。これを対象としたパターンマッチにより、節境界の位置を網羅的に検出し、節の種類を特定するシステムを構築している。自動検出した節とその種類を人手により確認し修正を施すことで、CBAP は 97% 以上の非常に高い精度での節境界の網羅的検出に成功している。また、丸山他（2004）では、『日本語話し言葉コーパス』に CBAP を適用し、句点を含まない音声コーパスの発話にも応用できることを示し、CBAP による節境界検出の有用性を述べている。当時、節境界の網羅的な検出や検出した節の種類の特定についての研究は少なく、丸山他（2004）に見られるような日本語学と工学の両分野の協働によるプログラム開発は先駆的、画期的であったと言える。

佐藤・丸山（2015）は、節境界認定システムである Rainbow3 の内容について報告している。「CBAP は、特定の形態素解析システム（ChaSen/IPAdic）に依存していることおよび、形態素解析結果の文字列を書き換える方式で実装されているため、保守性・拡張性に難がある」（佐藤・丸山 2015: 225）ことから、従来のシステムの改良ではなく、新システムの作成に至ったとしている。佐藤・丸山（2015）による Rainbow3 は 3 つのステップで節境界を認定するというシステムである。1 つめのステップで文節境界を認定し、2 つめのステップでは前段階で境界を認定した各文節について属性の認否を行う。文節の属性とは、述語文節か否か、助動詞文節か否か、連体修飾の可否、節末機能文節か否かの 4 つである。最後のステップで、先行の 2 つのステップでの認否判断のみを判断基準として、節境界の認否を決定するというものである。佐藤・丸山（2015）では、節境界認定のシステムの 2 つめのステップである 4 つの文節の属性のうち述語文節の認定と節末機能文節の認定について詳細な検討を行っている。

さらに、佐藤他（2016）は、標準的な節境界を定めることを目的とした節境界の付与システムの開発・改良への取り組みを説明し、Rainbow4 の節末境界の付与規則について説明している。詳細は定かではないが、基本的には Rainbow3 で用いられた節認定システムに改良を加え、BCCWJ に対する節認定付与の結果から、節認定規則についての検討と調整を重ねていることが述べられている。判定詞や終助詞の扱いおよび形容詞連体節の認定の難しさ、節の種類の特定に意味的な要素が含まれる場合の節ラベル付与の難しさなど、残された課題を明らかにしつつ、「標

準的な節境界を定める」ための取り組みが進んでいることを報告している。

丸山他（2016）は、前述した佐藤・丸山（2015）、佐藤他（2016）と同様に Rainbow システムについて述べている。佐藤・丸山（2015）、佐藤他（2016）が主として工学的視点から節境界の検出について述べているのに対し、丸山他（2016）では主として日本語学的視点から節境界ラベルの分類について、共同研究の成果を報告している。

丸山他（2016）では、日本語テキストに節境界ラベルを自動的に付与する節境界解析システム作成にあたり採用した、節の分類体系と節境界ラベルの分類体系について整理して述べている。日本語文法研究における従属節の分類に関する諸説を概観して比較したうえで、Rainbow システムの分類体系では、「全体の枠組みとしては、基本的に益岡・田窪（1992）を踏襲しているが、従属節を大きく3つに分けて、並列節を連用節の下位に位置づけ」（丸山他 2016: 1114）としている。節境界ラベルの分類体系は3段階からなっており、3つの従属節の下には、意味・機能による分類が設けられ、さらにその下位は多様な節末形式により分類されている。

また、丸山他（2016）では、日本語学の視点から、節境界解析の際、節境界認定や節境界ラベルの付与にあたり問題となる表現について例を挙げて指摘している。節境界を認定する位置により連用節になる場合と連体節になる場合があること、形容詞により名詞を修飾する連体節の節認定や、部分並列や述語が省略された構文の節認定は、可否判断が難しいこと、文脈の解釈により節境界ラベルの特定が難しい同形の節末形式が存在することなどについて検討し、節境界をめぐる多様な問題が存在することを示している。

ここまで日本語従属節の分類に関して、本研究を進めるにあたり多くの知見を得た最新の関連研究について整理して述べた。以下では、関連研究で明らかになったこと、指摘されていることを踏まえつつ、本稿の核心である、日本語従属節への「鳥バンク」基準のアノテーション作業でのタグ付与の齟齬と「鳥バンク」節間意味分類体系の関連について議論する。

3. 問題と修正

本節では、まず BCCWJ の一部に対する「鳥バンク」基準のアノテーション作業（1次作業）について、松本他（2017）で指摘した問題を整理して述べる（3.1 節）。次に、本研究で着目する意味分類のタグ付与の齟齬に絞り、1次作業の作業仲間齟齬と2次作業・3次作業において同一作業者が行った作業間齟齬を併せて再検討した結果、最終的に残った齟齬について述べる（3.2 節）。意味分類アノテーション作業の最終段階で残存した意味分類タグ付与の齟齬について、「鳥バンク」節間意味分類体系に関連する問題を意味分類階層、補足節、名詞修飾節、副詞節、並列節、「節間キーワード」の6項目に分け、項目ごとに事例を挙げて議論したうえで修正提案を行う。

なお、本稿ではアノテーション作業で付与されたタグを《第1段階（第2段階）（（第3段階））》のように示し、いずれかの段階の分類タグのみを示す場合は《第1段階》、《（第2段階）》、《（（第3段階））》のように表記する。付与タグの齟齬を示す場合は《第1段階（第2段階）（（第3段階））》－《第1段階（第2段階）（（第3段階））》と表す。

3.1 「鳥バンク」基準のアノテーション作業（1次作業）における問題

「鳥バンク」基準のアノテーション作業の概要を示す。まず、「鳥バンク」の「意味類型パターン辞書」のうち節境界パターンを UniDic 品詞体系に対応させた節自動解析器により、可能な節境界を全て列挙⁷する。この可能な候補を見ながら、作業員2名⁸（作業員A、作業員B）がお互いのアノテーションを見ずに1次タグ付与作業（以下、1次作業とする）を行う。その後、作業員Aが節自動解析器の出力および作業員A、Bの1次作業結果を見ながら2次タグ付与作業（以下、2次作業とする）を行う。作業対象はBCCWJ コアデータ⁹のうち出版・新聞の一部の54ファイル（優先順位00001～00054:A集合）とする。

作業では、節を「複文を構成するところの、述語を中心とした各まとまり」（益岡・田窪1992:4）と定義したうえで、述語を含む2文節¹⁰以上からなる節に意味分類タグを付与した。また、従属節の意味分類タグは、「鳥バンク」節間意味分類体系のみを手掛かりに、節末の接続表現の右端に付与することとする。

松本他（2017）では、1次作業で生じた作業員間の意味分類タグ付与の齟齬を日本語学の視点から分析し、従属節の意味分類アノテーション1次作業の問題点について以下の通り指摘している。

- i) 作業員間のタグ付与齟齬の約57%が節認定において生じ、表3に示す通り、その77%が名詞修飾節と副詞節において生じた。節認定で生じた齟齬とは、作業員の一方は意味分類のタグを付与したが、もう一方は節認定を行わなかったために、意味分類タグを付与しなかったことにより生じた齟齬である。

表3 節認定の齟齬の頻度

補足節	名詞修飾節	副詞節	並列節
102 (16%)	275 (44%)	207 (33%)	47 (7%)

※（ ）内は、節認定の齟齬が生じた総頻度に対する、各従属節の節認定の齟齬頻度の割合

節認定の問題に関しては、①何を節と認定するか ②どこからどこまでを節と認定するかという両面から論じる必要がある。述語の性質、意味から節を認定することは極めて難しく、厳密に節認定を行おうとすれば、認定はますます複雑化することが予想される。節認定に関して、文法上の議論の余地がある事例においても、広めの許容範囲を認め、場合に

⁷ https://github.com/masayu-a/clause_pattern（2017年9月7日確認。現在は公開されていない）

⁸ 日本語母語話者であるが、日本語の研究者等ではない作業員2名によって作業が開始された。作業員にはアノテーションについてのトレーニング等は施されておらず、「鳥バンク」節間意味分類体系だけを手掛かりにアノテーション作業をした。

⁹ 高精度の解析を実現するために整備された「形態素解析器・長単位解析器の学習用データ」であり、「短単位・長単位とも自動解析後に、全データに対して人手作業による確認を行い、誤解析の修正を行った」ものである。コアデータはBCCWJ全体の約1%であり、6つのレジスターにより構成されている。各レジスターの規模は、出版サブコーパスの新聞（204,050）、雑誌（202,268）、書籍（308,504）、および特定目的サブコーパスの白書（197,011）、ウェブのYahoo!知恵袋（93,932）、Yahoo!ブログ（92,746）である（（ ）内は延べ語数）。（小椋2014: 81）

¹⁰ 文節の定義は、小椋他（2011）の文節認定規程に従う。

よっては複数の節情報タグを付与することも検討し、柔軟に対応できる意味分類基準策定の可能性を探ることが求められる。

- ii) 従属節の意味分類¹¹の齟齬については、機能と意味に分けて分析する。作業により従属節に付与したタグ（第1段階4種類）の種類が異なるという、従属節の機能における分類タグ付与の齟齬の頻度は表4の通りである。表の薄墨色部分は、例えば《補足節（名詞節）》－《補足節（疑問節）》などのように、第1段階ではタグ付与の齟齬がなく、第2段階以下に齟齬が生じたものの頻度を表している。

表4 従属節の機能分類における作業者A,B間齟齬の頻度（1次作業）

A \ B	補足節	名詞修飾節	副詞節	並列節
補足節	8	4	7	8
名詞修飾節	11	179	4	0
副詞節	12	6	85	125
並列節	1	0	26	2

従属節の機能分類で生じた齟齬は《並列節》－《副詞節》間で最も多く、従属節機能分類の齟齬の74%を占めていた。また第1段階（大分類）の機能分類は一致したが、第2段階以下で齟齬が生じたものも認められた。これは、例えば作業者の一方が《名詞修飾節（内容節）》タグを付与し、もう一方が《名詞修飾節（補足語修飾節）》タグを付与したような場合である。この種の齟齬は、名詞修飾節の下位で最も多く生じた。名詞修飾節の下位での齟齬の75%は、《（内容節）》－《（補足語修飾節）》－《（縮約形修飾節）》間および補足語修飾節のさらに下位の《（（限定的））》－《（（非限定的））》間で生じている。

3.2 意味分類タグ付与の齟齬と「鳥バンク」節間意味分類体系の問題および修正提案

本節では、松本他（2017）で明らかにした「鳥バンク」基準のアノテーション作業の2つの問題のうち、節に付与された意味分類タグの齟齬に着目した議論を行う。1次作業での齟齬に、2次作業と3次作業での齟齬を併せて再検討し、最終的に残存したタグ付与齟齬がアノテーション作業で基準とした「鳥バンク」節間意味分類体系に起因していることについて述べる。タグの誤入力、誤分類を除外し、最終作業で確認された、従属節の機能レベルで生じたタグ付与の齟齬の頻度は表5の通りである。表4と同様に、無色のところは第1段階で齟齬が生じた頻度を表し、薄墨色のところは第2段階以下で齟齬が生じた頻度を表している。ただし、表5では、《補足節（名詞節）》－《名詞修飾節（内容節）》というタグ付与齟齬と《名詞修飾節（内容節）》－補足節（名詞節）というタグ付与齟齬の区別を行わず、タグ付与齟齬が生じた2つの節に着目して集計した頻度を示している。

¹¹ 「鳥バンク」では、従属節の機能と意味を厳密に分類していない。本稿では、鳥バンクの第1段階の分類を従属節の機能における分類と解釈した。

表5 再検討後の従属節機能分類における齟齬の頻度

	補足節	名詞修飾節	副詞節	並列節
補足節	0	6	5	0
名詞修飾節		10	0	0
副詞節			42	88
並列節				1

最終作業で残った 152 例のタグ付与の齟齬のうち、「鳥バンク」節間意味分類体系の第 1 段階(大分類)である、従属節の機能レベルの分類で生じた齟齬は 99 例(無色部分の合計)、第 2 段階以下の、主として意味レベルの分類において生じた齟齬は 53 例(薄墨色部分の合計)であった。

以下に、「鳥バンク」節間意味分類体系が、アノテーション作業のタグ付与の齟齬にどのような関連しているか、齟齬が生じた段階(第 1 段階と第 2 段階以下)ごとに、また第 2 段階以下では従属節の種類ごとに事例を挙げながら詳述する。そして「鳥バンク」節間意味分類体系を日本語従属節に適応する意味分類基準に再構築するための修正提案を行う。

事例はタグ付与に齟齬のあった従属節を含む文を示し、アノテーションした語に下線を付した。長すぎる文については文脈判断に支障がない範囲で前後を省略した。事例にはサンプル ID と作業ファイル名¹²を記した。

3.2.1 従属節の機能分類における齟齬と分類階層の修正

まず、従属節の機能レベル(第 1 段階)で生じたタグ付与の齟齬について述べる。従属節の機能レベルの齟齬とは、アノテーション作業において一方が補足節、他方が副詞節とタグ付与したような、「鳥バンク」節間意味分類体系の第 1 段階において生じた齟齬である。付与タグの再検討の結果、従属節の機能レベルに残存した齟齬は《補足節》－《名詞修飾節》,《補足節》－《副詞節》, および《副詞節》－《並列節》で生じていることが明らかになった。

「鳥バンク」節間意味分類体系では、従属節の節末に現われる接続の表層形式を「節間キーワード」としている。従属節へのアノテーション作業では、この「節間キーワード」を手掛かりに意味分類のタグ付与を行った。2.1 節(の表 2 およびその説明)で述べた通り、補足節では「節間キーワード」の形式名詞や引用符などが標識となるため齟齬が生じにくいと考えられるが、その標識である「節間キーワード」に起因して、タグ付与の齟齬が生じた例が確認されている。

以下に従属節の機能レベルで生じたタグ付与の齟齬について、《補足節》－《名詞修飾節》,《補足節》－《副詞節》,《副詞節》－《並列節》の順に事例を挙げて、「鳥バンク」節間意味分類体系との関連で議論したうえで、修正提案を行う。

《補足節》－《名詞修飾節》

- (1) フセイン政権が崩壊したことが重要で、大量破壊兵器開発の有無の確認も重要だが、根本的な問題ではない」との見解を示して [PN3d_00002, file00026]

¹² 作業ファイルは <https://github.com/masayu-a/BCCWJ-ANNOTATION-ORDER> 参照。

表5の《補足節》－《名詞修飾節》の齟齬にあたる6例全てが「との」の分類において生じていた。一方は「と」までを引用節と考え、《名詞修飾節（その他）》の「節間キーワード」である「節＋の＋名詞」を手掛かりに《名詞修飾節》とタグ付与した。他方は引用符と「と」を手掛かりに《補足節》とタグ付与したため、齟齬が生じたものである。「との」に関しては、益岡・田窪（1992: 204）が連体節の下位分類である内容節について「引用が関係する名詞は、内容節を伴う場合「という」, 「との」を伴う。」と述べているが、「鳥バンク」節間意味分類体系の《内容節》の「節間キーワード」に「との」は含まれていない。

《補足節》－《副詞節》

- (2) CBS テレビでは「大量破壊兵器に関する多くの文書を入手しており、多くのフセイン政権関係者の取り調べも続けている」として、搜索予定の関連施設がまだ多数あることを明らかにした。[PN3d_00002, file00026]

《補足節》－《副詞節》で生じた齟齬の5例全てが、事例（2）と同様の「として」の意味分類で生じた齟齬であった。事例（2）では、一方は引用符と引用節の「節間キーワード」である「と」から《補足節》のタグを付与し、他方は「として」が副詞節（判断）の「節間キーワード」であることから《副詞節》のタグを付与したことで齟齬が生じている。池原（2009）では、「として」を「節間キーワード」とした副詞節（判断）の文例として「契約の不履行にあたるとして解約を申し出る。」を示している。「鳥バンク」節間意味分類体系では、副詞節（判断）と補足節（とくに引用節の間接引用）の「節間キーワード」には、この他にもいくつかの重複が見られるが、このような「節間キーワード」の重複は意味分類タグ付与の齟齬の要因となり得る。

《副詞節》－《並列節》

- (3) 仏政府には債務問題を財政面だけでなく、主権回復など政治プロセスにリンクさせ、イラク再建を国際社会全体で取り組む契機にしたい意向が働く。[PN3c_00002, file00025]
- (4) 昨年十月から二カ月間、アジア六カ国・地域を回り、中国で出会った無名の日本人ミュージシャンの話に驚いた。[PN4g_00002, file00029]
- (5) 「三遊亭王楽の天狗道場・銀座編～円楽と東西の若手達」と題して、三遊亭円楽と王楽のほか、桂つく枝、柳家三三が出演する。[PN4g_00002, file00030]
- (6) 研究会はPTA 広報紙や学校新聞などを作る際の問い合わせにも応じ、情報を提供している。[PN5a_00002, file00031]

表5で示した通り、再検討後の従属節の機能レベルの分類における齟齬では、《副詞節》－《並列節》の齟齬が最も多く、機能レベルの齟齬の9割を占めている。1次作業において作業員Bは（3）～（6）を《並列節（順接的並列）》としているが、作業員Aは（3）を《副詞節（手段）》、（5）を《副詞節（付帯状況）》としている。（4）、（6）については《副詞節（手段）》としながらも、《副詞節（付帯状況）》という解釈もありうるという注記を残している。最終作業においてもこの齟齬は解消されず、残存している。《副詞節》－《並列節》の齟齬の場合、副詞節の下位で、手段・

付帯状況・因果関係のいずれを選択するかで揺れを示し、(4)、(6)のように別のタグを欄外に注記しているものも確認された。

益岡・田窪(1992: 208)は、総記の並列が述語の連用形とテ形のいずれによっても表現されることを述べ、「テ形は、並列の関係を表現する用法に加えて、述語を修飾する用法、すなわち、副詞節を作る用法を持つ。副詞節として働く場合、具体的には、原因、手段、付帯状況、等を表す。」としている。つまり、(3)～(6)の事例は、副詞節と並列節の2つの用法を共有していると言うことができる。このように従属節の意味に複数の解釈が可能な場合は、両論併記を認めるアノテーション基準を定めるべきであるが、従属節意味分類のアノテーション基準については稿を改めることにする。

「鳥バンク」基準のアノテーション作業の機能レベルで生じたタグ付与の齟齬の分析に基づいて、以下の通り「鳥バンク」節間意味分類体系の修正を提案する。

補足節と名詞修飾節、副詞節における齟齬に関しては、前述した通り限られた場合にのみ生じている。それが「節間キーワード」に起因していることが明らかであるため、「節間キーワード」の整備が必須である。他にも「節間キーワード」が齟齬の原因となった事例が確認されているが、「節間キーワード」の修正については、改めて3.2.6で詳しく述べる。

機能レベルでの齟齬が最多である副詞節・並列節に関しては、日本語学の視点からいずれの用法も認められる従属節に択一的にタグを付与する妥当性は高いとは言えない。丸山他(2016: 1114)では、日本語文法の先行研究における従属節の分類を比較し、「英語などの言語では、副詞節と並列節(等位節)は形態的に大きく異なるため、両者を区分する必然性があるが、日本語の並列節の場合、(中略)形態的には連用接続形式の一種とみなしてよい。」としている。本研究では、丸山他(2016)に倣い、「鳥バンク」節間意味分類体系の副詞節と並列節を連用節の下位に分類し、従属節の第1段階での分類を補足節・連体節・連用節の3分類と修正することを提案する。「鳥バンク」節間意味分類体系の再構築によって、本研究で提案する従属節の意味分類基準の構成は表6の通りである。

表6 修正案における意味分類基準の構成

段階	分類の着目点	分類項目数
第1段階	統語的特徴	3
第2段階	機能	9
第3段階	意味	20
第4段階	形式	28

「鳥バンク」節間意味分類体系では、従属節の意味・機能を4段階に分類しているが、第2段階以下に意味、機能、形式の混在が見られる。これを第1段階は統語的特徴、第2段階は機能、第3段階は意味、第4段階は形式による分類となるよう整理する。「鳥バンク」節間意味分類体系は英語-日本語機械翻訳のための体系であるため、翻訳の際、同じパターンに当てはまらない表現については、分類項目の新設・細分化を行っている。本研究では、日本語従属節のアノテ-

ションの精度を保つための意味分類基準の策定を目指すため、分類項目は同義のものを可能な限り集約し、浅いレベルでの最小限の分類にとどめる。なお、意味・機能をどの深さまで分類するかという議論については今後の課題とする。

3.2.2 補足節の下位分類における齟齬と分類項目の修正

補足節の下位分類に残存したタグ付与の齟齬は認められていない¹³が、「鳥バンク」節間意味分類体系の再構築を行うにあたり、日本語学の視点から若干の修正提案を行う。補足節の分類項目について表7に「鳥バンク」節間意味分類体系（左）と修正案（右）を併記して示した。

表7 補足節の意味分類体系の修正案（「鳥バンク」との比較）

「鳥バンク」節間意味分類			
大	中	小分類 (1)	小分類 (2)
補足節 (補語相当節)			
	名詞節		
	コト型	形式・意味により 8 項目に細分類	
	ノ型		
		強調構文	
	トコロ型	形式・意味により 3 項目に細分類	
	節＋格助詞型		
疑問節			
	選択疑問文		
	疑問語疑問文		
引用節			
	直接引用		
	間接引用	ト型	
		トハ型	
		ヨウニ型	
		ナドト型	
		ナンテ型	
		選択疑問節型	
		疑問語疑問節型	

「鳥バンク」節間意味分類の修正案				
第 1 段階	第 2 段階	第 3 段階	第 4 段階	
補足節				
	名詞節			
			コト＋格助詞等	
			ノ＋格助詞等	
			トコロ＋格助詞等	
			節＋格助詞等	
	疑問節			
			選択疑問文	
			疑問語疑問文	
	引用節			
			直接引用	
		間接引用・ト		
		間接引用・ヨウニ		
		間接引用・ナドト		
		間接引用・ナンテ		

補足節に関しては、最終作業において全ての齟齬が解消されたが、1次作業で生じた齟齬や、補足節の認否に関する日本語学上の議論を参考に、分類項目の集約などについて検討すべき課題を以下に挙げる。

① 形式名詞が述語位置に現われる場合

「鳥バンク」節間意味分類体系では補足節の下位の名詞節において、形式名詞が述語位置に現われる場合（日本語の文例「来年からそれは廃止になるとの事だ」（下線は筆者による）など）も《補足節（名詞節）》に分類しているが、本研究では、これを《補足節（名詞節）》に分類しない立場をとる。「鳥バンク」節間意味分類体系では、補足節の下位の名詞節を「節＋形式名詞」

¹³ 修正提案を行った箇所に節認定の問題を含んでいるものもあるが、本稿では節認定の問題を扱わない。

の形で名詞相当の働きを持ち、格助詞を伴って主節の述部を補う要素となる節」としている。この定義に基づいて検討すると、「廃止になるとの」は形式名詞「事」の名詞修飾節であるので、「節＋形式名詞」の形で名詞相当の働きを持つ」という名詞節の定義に合致している。しかし「格助詞を伴って主節の述部を補う要素となる節」という定義には合致していないため、本研究では「ことだ」「のだ」など形式名詞が述語位置に現われる場合については《補足節（名詞節）》に分類しないとの修正を行う。

「鳥バンク」節間意味分類体系では、上記のように「こと（事）」が「との」に後接している場合も、述語の基本形、タ形、連体形に後接している場合も、述語位置に現われる形式名詞「こと」「の」「ところ」を全て補足節に分類している。述語位置に現われる「こと」「の」「ところ」については諸説がある。益岡・田窪（1992: 29）は、「のだ」「ことだ」など「形式名詞＋「だ」」を「述語（動詞、形容詞、（名詞＋）判定詞）の基本形、タ形、連体形に接続して複雑な述語を作る語を「助動詞」に分類している。日本語記述文法研究会（2008: 13）は、「述語に対して主語や補語の関係にある節を補足節という。「だ」「である」などを伴って述語化することもある。」として補足節とし、その下位の分類では「末尾に形式名詞「こと」「の」「ところ」を伴って名詞と同様の性質をもつ節を名詞節という。文において主語や補語として用いられったり、述語になったりする。」（日本語記述文法研究会 2008: 15）として名詞節に分類している。この問題については議論の余地があることを認めたくて、本研究では格助詞を伴わない述語位置の形式名詞については全て、「鳥バンク」節間意味分類体系の定義の文言通りに判断し、補足節には分類しない。形式名詞を修飾する節（上記の日本語の文例については、「廃止になるとの」）を《名詞修飾節（形式名詞修飾節）》（後述の表 8 参照）に分類することとする。

② 第 4 段階の選択疑問節、疑問語疑問節の分類

「鳥バンク」節間意味分類体系では、名詞節のコト型、トコロ型、および引用節の下位に選択疑問節、疑問語疑問節を含む形式を区別して分類する項目が設けられているが、これは英語－日本語対訳のための分類であると考えられる。日本語だけを分類する場合、名詞節、引用節の下位での平叙文、疑問文の区別は特段に意味を持たないため、修正案ではこの分類項目を設けない。

また、益岡（1997: 12）は、引用節に関して、「補足語の働きをするものとみられるにもかかわらず、名詞の性格は有していない」と述べている。「引用節の処遇を巡って慎重な検討が必要」であるとしたうえで、主節を修飾する連用節とみなすことができる引用節があることを重視し、引用節を連用節として扱うという立場をとっている。このように引用節の分類についても議論の余地があることを承知したうえで、本研究では「と」「ように」「などと」「なんて」に導かれる節を《引用節》に分類する。

③ 格助詞を格助詞等に拡大

形式名詞「コト」「ノ」「トコロ」を伴うものとして、格助詞以外にも益岡・田窪（1992: 50）による提題助詞、取り立て助詞や、小椋他（2011）に「複合辞・助詞相当句」と整理されているもののうち、「助詞－格助詞」に品詞分類されている複合辞（以下、格助詞相当句とする）も含め、述語を補う要素となる節を《補足節（名詞節）》に分類することとする。本稿では「コト」「ノ」

「トコロ」に伴う格助詞，提題助詞，取り立て助詞，格助詞相当句を格助詞等とする。

3.2.3 名詞修飾節の下位分類における齟齬と分類項目の修正

第1段階の分類において付与したタグは一致したが，下位の第2段階で付与したタグに齟齬が生じた名詞修飾節について，事例を示しながらタグ付与の齟齬の原因について議論し，分類項目の修正提案を行う。

名詞修飾節の下位分類での齟齬は10例であり，第2段階以下での齟齬の2割に満たない。その10例中8例は，第2段階の《(内容節)》－《(縮約形修飾節)》で生じている。「鳥バンク」節間意味分類体系では，内容節を「被修飾名詞の内容を表す節。被修飾名詞と修飾節が同格の関係にあるタイプ」と，縮約形修飾節を「被修飾名詞と修飾節が格関係にも，同格関係にもなく，意味的に間接的な関係にあると考えられる節。修飾節と被修飾名詞の間に何らかの説明が省略されていると考えられるタイプ」と説明している。名詞修飾節の下位の《(内容節)》と《(縮約形修飾節)》で齟齬のあった例は以下の通りである。

《(内容節)》－《(縮約形修飾節)》

- (7) 不用なごみを減らす社会づくりに貢献する中小企業への対策が急がれる。[PN4g_00002, file00030]

事例(7)では，「不用なごみを減らす(ことを目指す)社会づくり」と解釈すると《(縮約形修飾節)》とタグ付与することも可能であり，「不用なごみを減らす(という)社会づくり」として《(内容節)》に分類することも可能である。事例(7)は，名詞修飾節を内容節とするか縮約形修飾節とするか判断に迷う例である。

益岡・田窪(1992)では，被修飾名詞の種類により，内容節を3つに分けて説明してはいるが，それぞれに命名して分類することはせず，「鳥バンク」節間意味分類体系で縮約形修飾節とされているものも内容節の一部として扱っている。名詞修飾節を詳細に分類することの可否については議論の余地がある。

《(内容節)》－《(その他)》

- (8) 一九九七年に「六つの目しかない骰子を振りながら，七つ目を求めているような表現者でありたい」との願いを込めて発足したグループ。[PN4g_00002, file00029]

事例(8)は，一方が「との」を伴う引用部分を「願い」の《(内容節)》とタグ付与したのに対し，他方は「引用節+の+名詞」という形式から，名詞修飾節の第2段階の《(その他)》の「節間キーワード」である「節+の+名詞」を手掛かりに《(その他)》タグを付与した例である。これは「鳥バンク」節間意味分類体系を手掛かりにタグ付与したことで生じた齟齬であると言える。

この例に関連して，名詞修飾節の下位分類には，もう1点，別の問題がある。「鳥バンク」節間意味分類体系では，第2段階で「節間キーワード」を「節+の+名詞」とする《(その他)》とは別に《(用言+接続表現+の)》という分類項目も設けている。これは「[用言+接続表現+の]」

という形式で名詞を修飾する節」と説明され、日本語の文例には「これは次回へ備えての休息期間である。」などが挙げられている。この2種類の分類項目の区別は極めて不明瞭であり、従属節の意味分類に混乱をもたらすものである。

名詞修飾節に関する「鳥バンク」基準のアノテーション作業の齟齬の分析を基に、名詞修飾節の分類項目についての修正を提案する。表8に「鳥バンク」節間意味分類体系(左)と修正案(右)を併記して示した。

表8 名詞修飾節の意味分類体系の修正案(「鳥バンク」との比較)

「鳥バンク」節間意味分類			
大	中	小分類(1)	小分類(2)
名詞修飾節(連体節)			
	補足語修飾節		
		限定的	
		非限定的	
	内容節		発言・思考 事柄・その他
	縮約形修飾節		意味により7項目に細分類
	機能的表現		
		形式名詞修飾節	
		文末表現相当	
		慣用的表現	
		副詞節相当	
	用言+接続表現+		形式により3項目に細分類
	その他		

「鳥バンク」節間意味分類の修正案			
第1段階	第2段階	第3段階	第4段階
連体節			
	補足語修飾節		
	内容節	同格関係	
		非同格関係	
	形式名詞修飾節		
	その他		

名詞修飾節に関する第一の修正提案は、第1段階の名称を連体節とすることである。これは、後述する連用節の提案に伴うもので、連体節、連用節という対比的分類を示すためである(これ以降、連体節とする)。

被修飾名詞が連体節の述語の補語となる補足語修飾節の下位での限定・非限定の分類について述べる。連体節が限定的であるか、非限定的であるかの区別は文脈に依存すること、補足語修飾節がどの名詞を修飾するかという係先の被修飾名詞の特定に揺れが生じる場合、連体節の意味分類にも揺れが生じることから、本研究の修正では限定・非限定の分類を解消し、補足語修飾節には下位分類を設けないこととする。

《(内容節)》(《同格》)は、「鳥バンク」節間意味分類体系の《(内容節)》にあたるもので、連体節と被修飾名詞が同格関係にあるものである。《(内容節)》(《非同格》)は、「鳥バンク」節間意味分類体系の《(縮約形修飾節)》にあたるもので、連体節と被修飾名詞が同格関係にないものである。「鳥バンク」節間意味分類体系では《(縮約形修飾節)》の下位を意味により7項目に分類しているが、修正案では下位の細分類を設けない。

形式名詞を修飾する節は、分類基準も明確であるため、形式名詞修飾節として独立させる。被修飾語となる形式名詞を列挙し、注記しておくことが望ましい。また形式名詞の周辺には、前述したように格助詞を伴って補足節となるものや述語位置に現われるもの、「ため」のように形式名詞を含んで副詞節として働くと考えられるものなど、多種多様な表現用法が存在し、その現象

を巡って日本語学の視点での議論も尽きない。従属節アノテーションの精度を保つためには、これらの扱いについて事例ごとに規定を設け、改良していくことが求められる¹⁴。

3.2.4 副詞節の下位分類における齟齬と分類項目の修正

第1段階の機能レベルの分類において付与したタグは一致したが、下位の第2段階で付与したタグに齟齬が生じた副詞節について、事例を示しながらタグ付与の齟齬の原因について議論し、分類項目の修正提案を行う。

副詞節の第2段階以下に生じたタグ付与の齟齬は42例あり、《(手段)》と《(因果関係)》の間の齟齬が21例と最多であった。他には第2段階で生じた《(時)》と《(条件)》の間の齟齬や、第3段階で、因果関係の下位分類の《((原因))》と《((結果))》の間に生じた齟齬などがある。副詞節の下位分類において齟齬が生じたケースは、付与された2つのタグのうちの一方が《(因果関係)》であったケースが27例、《(条件・譲歩)》であったケースが5例、《(付帯状況・様態)》であったケースが5例見られた。この結果から、齟齬を生じる分類のバリエーションは少なく、限られた分類項目で齟齬が生じていることが明らかになった。以下にタグ付与の齟齬が生じた事例を挙げて、齟齬の原因について議論し、「鳥バンク」節間意味分類体系の修正を提案する。

《(因果関係)》－《(手段)》

- (9) 信号の約6万カ所を集中制御タイプなどに切り替え、約150万トンのCO₂が減らせるとする。[PN5a_00003, file00043]
- (10) 他派閥からも引き抜いて三十人から五十人の新派閥をつくることができるんだ。[PN2e_00002, file00021]
- (11) 東区のマスコットである「タッピー」をあしらい、親しみやすくした。[PN3e_00002, file00027]

《(因果関係)》－《(手段)》の齟齬が副詞節の下位分類での齟齬で最多の21例であった。「鳥バンク」節間意味分類体系では、《(手段)》を「主節の内容を実現するための手段を従属節で表す。」とし、《(因果関係)》を「従属節と主節で表される事態間の因果関係を表す。」としている。事例(9)では、この記事の著者が「約150万トンのCO₂」を減らすために「集中制御タイプなどに切り替え」ることを手段として積極的に用いることを強調したいと解釈すれば、《(手段)》とのタグが付与される。一方で、「集中制御タイプなどに切り替え」ることで「約150万トンのCO₂が減」という客観的事実を報告しているのであれば、《(因果関係)》のタグが付与されることも誤ったアノテーションではない。事例(10)(11)についても同様の解釈が可能であり、これらは書き手の意図が節の意味分類に影響を与える事例である。この問題については、丸山他(2016)も文脈(発話意図)により解釈が分かれる節が存在することを指摘しており、タグ付与の齟齬が生じることは避けがたいことであると言える。そもそも、従属節の内容が主節の内容を実現する手段である

¹⁴「コト」「ノ」「トコロ」+格助詞等については《補足節(名詞節)》のタグを付与するものとするが、「コト」「ノ」「トコロ」を修飾する連体節のアノテーションの要否は節認定基準とアノテーション基準に従うものとする。

ならば、主節と従属節に因果関係が生じるのは必須であることから、これらの下位分類の要否についても議論の余地がある。

《(時)》 - 《(条件・譲歩)》

(12) 青空の下で汗ばみながら、砂利を掘って設置し終わると、スピーカーで鳴き声を流す。
[PN5b_00002, file00031]

(13) このため、辞任が決まると、カード首席補佐官は米誌で「ローブ氏に対抗できる側近が必要だ」と述べ、勢力均衡の崩壊に懸念を示した。[PN2e_00004, file00052]

事例 (12) (13) は《(時)》と《(条件・譲歩)》の間で付与タグに齟齬が生じた例である。「鳥バンク」節間意味分類体系では、副詞節の第2段階の《(時)》と《(条件・譲歩)》のいずれにも「～と」を含む「節間キーワード」が複数存在する。

「鳥バンク」節間意味分類体系の分類項目の説明では、《(時)》は「従属節と主節が表す事態の時間的關係を表す。」となっており、下位は第4段階まで細分類されている。例えば「従属節の事態が起こると必ず主節の事態が起こることを表す。」と説明される《(時) ((事態の時)) (((必然・習慣)))》を見ると、「節間キーワード」には「と (必ず | いつも | きっと | 決まって)」があり、文例は「彼は興奮するといつも顔面が紅潮する。」と示されている。また「従属節の事態が起きた瞬間に主節の事態が起こることを表す。」と説明される《(時) ((事態の時)) (((瞬時)))》を見ると、「節間キーワード」が「と (直ぐ (に) | 間も無く | 途端 (に))」であり、文例「ベルがなるとすぐに芝居が始まった。」が挙げられている。

一方、「従属節と主節で表される事態間の依存關係を表す」と説明されている《(条件・譲歩)》も第4段階まで細分類されている。例としては、「既に成立した事態間の依存關係を表す。確定条件。」とされる《(条件・譲歩) ((偶有的)) (((個別的依存關係)))》の「節間キーワード」は「と となると」であり、文例は「いざ投票となると多くの共和黨員は民主党側に票を投じた。」である。また「現実から独立した、事態間の依存關係を表す。「もし」、「かりに」等を伴うこともある。」と説明される《(条件・譲歩) ((仮想的))》では、「節間キーワード」が「とすると」であり、文例としては「二つのことを同時にやろうとすると、結局虻蜂取らずになる。」が挙げられている。

これらは英語の when, if などの翻訳に関連する分類であり、日本語学的には副詞節の節末形式「と」は専ら条件節とみなすことが慣用となっている。益岡・田窪 (1992: 193) では、条件・譲歩を表す副詞節の説明の中で、「花子は、家に帰ると、すぐに友人に電話をかけた。」という文例を挙げて、「すでに成立した個別的事態についての依存關係を表すこともできる。」としている。

《((結果))》 - 《((原因))》

(14) あちこちで芝がはがれピッチ上に穴があくような状況は見られなかった。[PN2c_00002, file00020]

「鳥バンク」節間意味分類体系の因果關係の下位分類である《((結果))》と《((原因))》については、《((結果))》を「従属節の内容が起こった結果、主節の内容が起こることを表す。」とし、

《((原因))》を「従属節の内容が原因となって主節の内容が起こることを示す。」としている。事例 (9) ～ (11) の発話意図の解釈と同様に、事例 (14) も結果と原因のどちらを重視しているかという書き手の意図を汲み取って分類する必要がある例であると言える。この分類の要否についても議論の余地があると言えよう。「鳥バンク」節間意味分類体系で《(因果関係)》としている節を日本語学では下位分類せず原因・理由節として扱うことが慣用である。

副詞節に関する「鳥バンク」基準のアノテーション作業の齟齬の分析を基に、副詞節の分類項目についての修正を提案する。表9に「鳥バンク」節間意味分類体系（左）と修正案（右）を併記して示した。

表9 副詞節の意味分類体系の修正案（「鳥バンク」との比較）

「鳥バンク」節間意味分類			
大	中	小分類 (1)	小分類 (2)
副詞節 (副詞的連用節)			
時		事態の時	意味により 16 項目に細分類
		事態の継続期間	意味により 3 項目に細分類
因果関係		原因	意味により 9 項目に細分類
		結果	意味により 1 項目追加分類
		理由根拠	意味により 2 項目追加分類
条件・譲歩		法則的	意味により 1 項目追加分類
		偶有的	意味により 5 項目に細分類
		仮想的	意味により 1 項目追加分類
		反事実的	
		譲歩	意味により 3 項目に細分類
付帯状況・様態		付帯状況	意味により 3 項目追加分類
		様態	意味により 5 項目に細分類
逆接			意味により 9 項目に細分類
目的			意味により 8 項目に細分類
程度			意味により 12 項目に細分類
前提			意味により 2 項目に細分類
手段			意味により 5 項目に細分類
二者関係			意味により 7 項目に細分類
相関			意味により 2 項目に細分類
判断, 主観			意味により 6 項目に細分類
場面			意味により 6 項目に細分類
限定			意味により 3 項目に細分類
独立			意味により 6 項目に細分類
その他			
		助動詞相当表現	義務・不可避
		慣用的表現	意味により 4 項目に細分類
		副詞相当表現	
		発言・思考内容	
		2 文	

「鳥バンク」節間意味分類の修正案			
第 1 段階	第 2 段階	第 3 段階	第 4 段階
連用節			
	副詞節		
		時間	
		条件	
		譲歩	
		原因・理由・手段	
		付帯状況	
		様態	
		逆接	
		程度・比較	
		目的	
		その他	

修正案では、副詞節の分類項目を「鳥バンク」節間意味分類体系の第4段階での155種から第3段階10種に大幅に削減した。この分類項目数の差は、「鳥バンク」節間意味分類体系と本研究での修正案の目的の相違によるものである。「鳥バンク」節間意味分類体系は日本語－英語機械翻訳を目的とすることから、言語間の統語的、意味的な違いに対応するための粒度の細かい分類項目を設ける必要があったと思われる。しかし、本研究で提案する日本語従属節の意味分類基準は、日本語研究のためのものであり、従属節アノテーションの精度を保つためのものである。

そこで、副詞節の下位の分類は基本的に益岡・田窪（1992）に則り、許容範囲を広く取り、粗い粒度で分類できるように修正を加えた。また、項目名についても、可能な限り日本語学の慣用に従うように修正した。

具体的な修正点としては、3.2.1で述べた通り、副詞節を連用節の下位に分類した。また、齟齬の分析から、文脈や書き手の意図に強く依存することにより、意味分類のタグの特定に大きな揺れがあると判断した《(因果関係)》と《(手段)》については、第3段階において《((原因・理由・手段))》に集約し、その下位分類は設けないこととした。また、項目名は日本語学での慣用に従い、《(因果関係)》は《((原因・理由))》に、《(時)》は《((時間))》に改めた。

3.2.5 並列節の下位分類における齟齬と分類項目の修正

第1段階の分類において付与したタグは一致したが、下位の第2段階で付与したタグに齟齬が生じた並列節は、以下に挙げる事例が1件のみであった。タグ付与の齟齬について議論したうえで、分類項目の修正提案を行う。

《((総記の並列))》－《((例示の並列))》

- (15) 5人の女優が時にぶつかり、時にいたわり合う、その競演ぶりが見ものだ。[PN1b_00003, file00033]

事例(15)は、「～ぶつかり」を並列節と分類することに一致は見られたが、その下位で《((総記の並列))》のタグを付与するか、《((例示の並列))》のタグを付与するかでの齟齬が生じた例である。「鳥バンク」節間意味分類体系では、並列節を「従属節が主節に対して対等の関係で結びつくタイプ」とし、その下位の第2段階に《(順接的並列)》と《(逆接的並列)》を置き、《(順接的並列)》の下位を6分類している。事例(15)に付与されたタグはそのうちの1つで、《((総記の並列))》は「該当するものをすべて列挙する表現」、《((例示の並列))》は「該当するものの一部を例として列挙する表現」と説明されている。「節間キーワード」からのみ判断するのであれば《((総記の並列))》となるが、「節間キーワード」に「時に」という副詞は見当たらない。《((例示の並列))》の「節間キーワード」は「ば」、「たり」、「や」となっており、ここでも「時に」は見られない。益岡・田窪（1992）で事態が起こる頻度を表すアスペクトの副詞と説明されている「時に」や、文脈から考えると、《((例示の並列))》に分類するのが妥当であろう。

並列節に関する「鳥バンク」基準のアノテーション作業の齟齬の分析を基に、並列節の分類項目についての修正を提案する。表10に「鳥バンク」節間意味分類体系（左）と修正案（右）を

併記して示した。

表 10 並列節の意味分類体系の修正案（「鳥バンク」との比較）

「鳥バンク」節間意味分類			
大	中	小分類（1）	小分類（2）
並列節（並列型連用節）			
	順接的並列		
	総記の並列		
	例示の並列		
	累加の並列		
	平行		
	否定の並列		意味により 4 項目に細分類
	選択		
	逆接的並列		

「鳥バンク」節間意味分類の修正案			
第 1 段階	第 2 段階	第 3 段階	第 4 段階
連用節			
	並列節		
		順接的並列	
		逆接的並列	

並列節において最も問題となるのは、前述した通り、副詞節との関連である。2つの事態が互いに対等な関係で並列して記述されているが、意味的には並列節が原因、手段、付帯状況を表していると解釈できる例があることは 3.2.1 でも述べた通りである。また、《（順接的並列）》の下位 6 分類の可否については議論の余地がある。

そこで、本研究では、並列節を連用節の下位に分類すること、《（順接的並列）》は下位分類を行わないことを並列節についての修正提案とする。

3.2.6 「節間キーワード」の問題点と修正

「節間キーワード」については 2 種類の重複の修正を提案する。1つは複数の節に同形の「節間キーワード」が記載されている問題の修正であり、もう 1つは「節間キーワード」の中に別の節の「節間キーワード」が含まれている問題の修正である。

これまでに述べてきた齟齬の問題に関しても「節間キーワード」に起因しているものが含まれていたが、ここで改めて「鳥バンク」節間意味分類体系の日本語の文例を挙げながら、「節間キーワード」の問題点について議論し、修正提案を行う。以下に示す「鳥バンク」節間意味分類体系の日本語の文例のように、意味・機能の異なる節が同形の節末形式で表現されることは珍しくないため、混乱を招かないための記載の工夫が必要である。

なお、「鳥バンク」節間意味分類体系の日本語の文例を挙げる際、「節間キーワード」とされている語には波線を付した。

(16) 私は友達の顔色の悪いのに気付いた。《補足節（名詞節）（（ノ型））》

(17) 1 時には来るはずなのに友達はまだ来ない。《副詞節（逆接）》

(16) (17) の日本語の文例では、「節間キーワード」はいずれも「のに」であり、表層的には節末形式は同形である¹⁵。このような場合は、同形の「節間キーワード」が他に存在することを注記するなど、記載の方法を工夫する必要があると考える。

¹⁵ 節認定の問題が含まれるが、節認定については別稿に譲る。

「鳥バンク」基準のアノテーション作業において、同形の「節間キーワード」を持つ複数の従属節があることでタグ付与に齟齬が生じた例を挙げる。

《副詞節（条件・譲歩）》－《補足節（名詞節）》

- (18) 7月の交通量を比較調査したところ、12時間の通行車両は約7600台から約9100台に増えたが、[PN5a_00003, file00043]

事例(18)は1次作業で生じた齟齬である。「鳥バンク」節間意味分類体系では「節間キーワード」として、「ところ」を含む分類項目が複数存在している。事例(18)の従属節は格助詞を伴っていないことから補足節とするのは誤分類であるとの判断で修正されたため、2次作業以降は齟齬として抽出されていない。しかし、上記の齟齬のように「節間キーワード」の「ところ」を含む分類項目の間で揺れが生じる可能性を以下の例により示す。

- (19) 丁度夕食が終わったところに彼がやってきた。
 (20) 私が車を降りようとしているところにちょうど彼女がふらっと通りかかった。

「鳥バンク」節間意味分類体系によると、(19)は《(名詞節)》、(20)は《(時)》で「節間キーワード」の「ところ(に|へ)」が現われる文例として示されている。これらは第1段階から異なる節となっているが、これらの意味的な違いを説明することも、どちらに分類することが妥当かを判断することも難しい事例である。

本節において他の事例でも「節間キーワード」が複数の節で重複している問題を多々指摘したが、「節間キーワード」の重複に問題があるわけではない。重複していることを注記しておくことが重要である。

次に1つの「節間キーワード」に異なる従属節の「節間キーワード」が含まれている問題について「鳥バンク」節間意味分類体系の日本語の文例を挙げて述べる。

- (21) 一生懸命勉強しただけのことはあって、今度の試験はよくできた。

(21)は「鳥バンク」節間意味分類体系の《副詞節（因果関係）((原因))》の日本語の文例である。「節間キーワード」は下線部の「だけのことはあって」となっているが、この「節間キーワード」には「だけ」という副詞節（程度）の「節間キーワード」と、「節+の+名詞」という名詞修飾節の「節間キーワード」、節末の「て」という《副詞節（因果関係）((原因))》の「節間キーワード」が含まれている。

「できるだけことはした。」という文を見ると、「だけのことは」が《((原因))》を表現する「節間キーワード」たり得ないことがわかる。(21)を《副詞節（因果関係）((原因))》としているのは「あって」というテ形述語であり、節末の「て」により意味分類される。基準を設けずに、個別の表層形式を「節間キーワード」に加えると、節認定の問題も生じ、分類基準が複雑になり過ぎる危険がある。

(21)の例を修正案に則って分類すると、「だけのことはあって」という節末形式はなくなり、

「節+の」「ことは」「て」に分解されそれぞれの項目に分類される。「だけの」は節末形式「節+の」によって《連体節（その他）(((節+の)))》に、「ことは」は《補足節（名詞節）(((コト+格助詞等)))》に、「あって」の節末形式「て」は《連用節（副詞節）((原因・理由・手段))》に分類される。ただし、この全てに意味分類のタグ付与をするか否かは、どこまでを節とするかという節認定のことになるので、稿を改めて議論する。

以上の問題の分析から、アノテーションしたデータの信頼性を確保するためにも次の修正が求められる。「節間キーワード」の2種類の重複問題について提案する修正は以下の通りである。

- i) 意味分類を異にする複数の節に同形の「節間キーワード」が存在する場合は、その「節間キーワード」を持つ節を全て参照できるように注記する必要がある。
- ii) 「節間キーワード」の中に別の節の「節間キーワード」が含まれている場合、「節間キーワード」の中には、原則として、他の項目の「節間キーワード」を含むことを許容しないとす
る修正を提案する。

また、「節間キーワード」は意味分類の手掛かりとして非常に有効であるが、「鳥バンク」節間意味分類体系の「節間キーワード」は節末表現と、節外の語（節に後接する語、共起する副詞、主節の文末表現など）が混在しているため、以下の修正を加える。

- iii) 「節間キーワード」を節末に限定する（節外に共起する語がある場合は、別途注記する）。

4. 日本語研究のための「鳥バンク」節間意味分類体系再構築の提案

本節では、3節で議論した「鳥バンク」節間意味分類体系を日本語従属節の意味分類に転用することで生じる問題と、それを踏まえた修正案を改めて整理する。3節では従属節の種類ごとに修正案を「鳥バンク」節間意味分類体系と併記して示したが、表11に本研究で提案する「鳥バンク」節間意味分類体系の再構築による日本語従属節意味分類基準の全容を、それぞれの分類項目の節末形式の例を加えて再掲する。

表 11 日本語従属節意味分類体系（「鳥バンク」の再構築）

統語	機能	意味	形式	節末形式
第 1 段階	第 2 段階	第 3 段階	第 4 段階	
補足節	名詞節		コト＋格助詞等	ことを・ことが・ことに…
			ノ＋格助詞等	のを・のが・のと・ので…
			トコロ＋格助詞等	ところを・ところで…
			節＋格助詞等	節＋に・節＋が・節＋を…
	疑問節		選択疑問文	か・や・やら・かしら…
			疑問語疑問文	疑問語＋か
	引用節		直接引用	引用符＋と
			間接引用・ト	と
			間接引用・ヨウニ	よう（に）
			間接引用・ナドト	など（と）
連体節	補足語修飾節			
	内容節	同格関係		（という）（との）
		非同格関係		
	形式名詞修飾節			
	その他		節＋の	節＋の
連用節	副詞節	時間		とき・ころ・際…
		条件		と・ば・たら…
		譲歩		ても…
		原因・理由・手段		ので・ため・から…
		付帯状況		ながら・つつ…
		様態		ように・ふうに…
		逆接		が・けれど…
		程度・比較		くらい・ほど・だけ…
		目的		ために・ようと…
		その他		
	並列節	順接的並列		連用中止・て…
		逆接的並列		が・けれど（も）…

5. おわりに

「鳥バンク」のホームページには、冒頭ページに「鳥バンクは、日本語表現意味辞書などの知的成果を広く流通・利用されて開かれた言語コミュニティの形成・拡大・発展に資し、以って言語文化等の発展に寄与するために、（中略）提供される知的財産です。」（池原 2007a）とある。本研究は、日本語研究の発展にとっても非常に価値の高い知的財産である「鳥バンク」節間意味分類体系を再構築し、アノテーション精度を保つための日本語従属節の意味分類基準を策定することについて検討した。BCCWJ（一部）の日本語従属節に対する「鳥バンク」基準のアノテーショ

ン作業において生じた意味分類タグ付与の齟齬の分析から、英語－日本語機械翻訳のための「鳥バンク」節間意味分類体系を日本語従属節の意味分類に転用することで生じる問題を明らかにし、「鳥バンク」節間意味分類体系の修正による日本語研究のための日本語従属節意味分類基準（「鳥バンク」節間意味分類体系の再構築）を提案した。

本稿では、まず本研究の基となった「鳥バンク」の開発研究と、本研究を進めるにあたり数々の知見を得た、日本語学分野と工学分野の協働による節境界認定についての最新の関連研究について述べた（2 節）。次に「鳥バンク」基準のアノテーション作業の問題点を整理した松本他（2017）で得たデータに基づき、意味分類タグの付与に齟齬があった事例と「鳥バンク」節間意味分類体系の関連について議論し、「鳥バンク」節間意味分類体系の修正を提案した（3 節）。それを踏まえ改めて、日本語研究のための日本語従属節意味分類基準案（「鳥バンク」節間意味分類体系の再構築）の全容を示した（4 節）。

本稿で扱った齟齬は、大きく 2 種類に分けることができる。1 つは「鳥バンク」節間意味分類体系の「節間キーワード」の重複などによるタグ付与の齟齬であり、もう 1 つは従属節の意味の曖昧性、解釈の揺れから生じたタグ付与の齟齬である。前者の齟齬の解消については意味分類基準の策定が求められ、後者の齟齬の解消については、アノテーション基準の策定が求められる。

本稿において、日本語従属節の意味分類基準策定のために検討した「鳥バンク」節間意味分類体系についての主な修正提案は、以下のようにまとめることができる。

- i) 第 1 段階においては、「鳥バンク」の補足節（補語相当節）・名詞修飾節（連体節）・副詞節（副詞的連用節）・並列節（並列型連用節）の 4 分類を、補足節・連体節・連用節の 3 分類に修正する。
- ii) 第 2 段階は機能、第 3 段階は意味、第 4 段階は形式に着目した分類を行う。
- iii) 「鳥バンク」節間意味分類体系の詳細な分類項目を日本語学の視点から見直し、大きな粒度で再分類する。
- iv) 同形の「節間キーワード」が複数の節の「節間キーワード」となる場合は、当該「節間キーワード」が複数存在することを注記し、参照できるようにする。「節間キーワード」の中には他の「節間キーワード」を含まないようにする。また、「節間キーワード」は節内に納まる節末形式とし、後接する節等に共起する語がある場合は別途注記する。

本稿では、タグ付与の齟齬が生じた節に着目して分析を行い、「鳥バンク」節間意味分類体系の再構築による日本語従属節意味分類基準の枠組みを提案したが、今後は分類項目ごとの節末形式、日本語の文例の再編成、注記の方法などを詳細に規定しなければならない。

本稿では、主として意味分類基準を策定することにより解消できる齟齬について論じ、意味の曖昧性、解釈の揺れから生じたタグ付与の齟齬の解消については論じなかったが、節の意味を 1 つに決めることが困難な場合があることを前提に、節の意味解釈を複数認め、両論併記を許容するアノテーション基準を整備しなければならない。また、何を節として、どこにアノテーションするかということについては、節認定という非常に大きな課題も残っている。アノテーションの

精度を保つためには、誰が、いつアノテーションしてもタグ付与に齟齬が生じない一定の基準を意味分類と作業の両面から策定することが必要である。

従属節について、南（1974: 130）は「問題の従属句がどの類に属するかは、必ずしも、テとかナガラ、バといった接続助詞のいかんによってあらかじめ決められるものではないということである。それは、その句を構成しているすべての要素およびその句の文中での文法的性格による。」と述べている。また、複文について、野田（2002:12）は「連用節と名詞節、連体節は、それぞれ形も働きも違うものであるが、その違いは本質的なものではない。」とし、節と文、文と文が集まったテキストの境界の曖昧さ（連続性）を指摘している。佐藤他（2016: 412）は節境界付与システムの設計において「ある方針に従って決めるしかない」というような問題が存在するとしている。池原（2009）のように、節間の意味分類は高度な判断を要するため、機械的な自動化が困難であり、多くの課題を解決しなければならないことを認めただうえで、人手による付与作業も決して容易ではないことを指摘し、人手付与作業を支援するために試行錯誤を繰り返し、意味分類を半自動化することに尽力している研究もある。上記の諸研究で述べられている通り、節境界や節の種類の特定が、あらゆるところに議論の余地と判断の難しさを有していることは間違いない。節認定や意味分類はその基準策定が非常に難しい問題であるだけに、今後の日本語研究の発展にとって、本研究の日本語従属節意味分類基準案のような叩き台が必要である。

参考文献

- 堀恵子・江田すみれ（2013）「web 公開予定文法用例検索システム「日本語文法項目用例文データベース『はごろも』」のレベル付けと学習者コーパスの比較」『第3回コーパス日本語ワークショップ予稿集』401-406.
- 池原悟（2007a）「鳥バンク（Tori-Bank）」<http://unicorn.ike.tottori-u.ac.jp/toribank>（2017年11月30日参照）.
- 池原悟（2007b）「意味類型パターン記述言語仕様書」<http://unicorn.ike.tottori-u.ac.jp/toribank/download/shiyou.pdf>（2017年11月30日参照）.
- 池原悟（2009）『非線形言語モデルによる自然言語処理』東京：岩波書店.
- Maekawa, Kikuo, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka and Yasuharu Den (2014) Balanced Corpus of Contemporary Written Japanese. *Language Resources and Evaluation (LRE48)*: 345-371.
- 丸山岳彦（2014）「現代日本語の多重的な節連鎖構造について—CSJとBCCWJを用いた分析—」石黒圭・橋本行洋（編）『話し言葉と書き言葉の接点』93-114. 東京：ひつじ書房.
- 丸山岳彦・柏岡秀紀・熊野正・田中英輝（2004）「日本語節境界検出プログラムCBAPの開発と評価」『自然言語処理』11(3): 39-68.
- 丸山岳彦・佐藤理史・夏目和子（2016）「現代日本語における節の分類体系について」『言語処理学会第22回年次大会発表論文集』1113-1116.
- 益岡隆志（1997）『複文』東京：くろしお出版.
- 益岡隆志・田窪行則（1992）『基礎日本語文法 改訂版』東京：くろしお出版.
- 松本理美・浅原正幸・有田節子（2017）「『現代日本語書き言葉均衡コーパス』に対する節の意味分類情報アノテーション—基準策定、仕様書作成の必要性について—」『言語資源活用ワークショップ2016 発表論文集』336-346.
- 南不二男（1974）『現代日本語の構造』東京：大修館書店.
- 日本語記述文法研究会（2008）『現代日本語文法6 複文』東京：くろしお出版.
- 野田尚史（2002）「第1章 単文・複文とテキスト」仁田義雄・益岡隆志（編）『日本語の文法4 複文と談話』1-62. 東京：岩波書店.
- 小椋秀樹（2014）「第4章 形態論情報」山崎誠（編）『書き言葉コーパス—設計と構築—』68-88. 東京：

朝倉書店.

小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕 (2011) 『『現代日本語書き言葉均衡コーパス』形態論情報規定集第4版 (上)』特定領域研究「日本語コーパス」平成22年度研究成果報告書.

佐藤理史・丸山岳彦 (2015) 「節境界認定に関する諸問題」『第8回コーパス日本語学ワークショップ予稿集』225–232.

佐藤理史・丸山岳彦・夏目和子 (2016) 「現代日本語書き言葉均衡コーパスに対する節境界付与」『言語処理学会第22回年次大会発表論文集』409–412.

Establishing Semantic Classification Rules for Japanese Subordinate Clauses: Reconstructing the Tori-Bank Semantic Classification System

MATSUMOTO Satomi

Doctoral Student, Ritsumeikan University / Project Collaborator, NINJAL

Abstract

This paper discusses the establishment of semantic classification rules for Japanese subordinate clauses based on a reconstruction of the Tori-Bank interclausal semantic classification system for Japanese to English machine translation, which is achieved by analyzing sentence patterns within Japanese-English corpora. As a conflict analysis of semantic classification tagging for subordinate clause annotation in the Balanced Corpus of Contemporary Written Japanese has revealed problems in the cross-application of this system to Japanese subordinate clauses, this paper attempts to reconstruct this classification system from a Japanese linguistics perspective. As context and intention affect the type of an adverbial clause, semantic classification rules can maintain the accuracy of clause annotation by proposing coarse-grained classification. The paper also summarizes the interclausal keyword problems inhibiting subordinate clause annotation (overlapping interclausal keywords between multiple clauses) and proposes improvements to reliability. While there are a number of challenging problems in clause boundary recognition and semantic classification of clauses, this research points to the potential for stable rules for the classification of Japanese subordinate clauses through reconstructing this system.

Key words: Tori-Bank, Japanese subordinate clause, semantic classification, Japanese clause classification annotation